DOCUMENT RESUME

ED 071 582 HE 003 717

AUTHOR Grasha, Anthony F.
TITLE Evaluating Teaching: Some Problems.
INSTITUTION Cincinnati Univ., Ohio. Inst. for Research and Training in Higher Education.
PUB DATE 72
NOTE 5p.; IRTHE, Vol. 4, No. 3 Spring 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *College Faculty; Evaluation Methods; *Evaluation Techniques; *Higher Education; *Professors; *Teacher Evaluation; Teacher Improvement; Teaching Quality

ABSTRACT

In the process of teacher evaluation, much needs to be taken into account other than a teacher's activities and behaviors inside the classroom. Also to be considered are the teacher's fixed goals for a particular class, his attitude toward his students, his peers' attitudes toward him, and an instructor's professional activities. Who should conduct the evaluation of teachers is something to be taken very seriously. Some possibilities are an independent evaluation agency, interested students, faculty and administrators in a college, or perhaps a committee composed of junior and senior members of a department. Who does it depends on what purpose the evaluation is to serve. Information derived from an initial evaluation should be used as feedback for the instructor in order that he might understand what others view as his faults. Then, a follow-up evaluation should be conducted to answer the following questions: (1) Did the instructor make an attempt to improve what he or she was doing? (2) Are the instructor's values with regard to things like professional research, writing and community service compatible with the institution's values? (3) Do colleagues perceive the instructor as a valuable member of the faculty? and (4) Have students reported that what they learned from the course, in class and out, helped them in decisions about what to major in, post-graduate study or on the job? The answers to these questions should help to determine whether the faculty member should be promoted or given tenure. (HS)

FILMED FROM BEST AVAILABLE COPY

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

OCT 11 1972

ED 071582

IRTHE

University of Cincinnati, Institute for Research and Training in Higher Education

Evaluating Teaching: Some Problems

There is now awareness of strong needs to implement methods conducive to the improvement of teaching and providing objective data for evaluation purposes. Faculty and administrators can no longer say merely that they recognize the problem and that they want to do something about it. Pressures from college administrations, students, legislatures, the job market, and other sources are demanding that action be taken. Current procedures range from informal student-initiated and -run evaluations to those having varying degrees of support from administrators and faculty. The goals of these procedures vary from telling students what the good and bad courses are to providing administrators with something more than hearsay information about who the effective teachers are.

Lost in discussions about how to gather the necessary data, and how reliable and valid student inputs can be, is an extremely important issue. Faculty members have hardly been overly enthusiastic towards evaluation systems. An important reason for this is that evaluation is an emotional issue and, consequently, defensive behaviors are evoked in members of the faculty. Most college teachers would acknowledge that teaching effectiveness needs to be improved, that objective methods must be established for eliminating the deadwood from their ranks and for making promotion and tenure decisions. But few are willing to admit that perhaps their own teaching needs to be improved, that they may be one of the instances of departmental deadwood, or that they have not earned a promotion. In a nutshell, evaluations are often seen as a threat to self-esteem and to job security. People have a tendency to resist such threats.

The resistance can take several forms, ranging from overt opposition by voting against a system to verbally criticizing it. K. E. Eble in *The Recognition and Evaluation of College Teaching* (November, 1970) has provided several familiar examples of the latter. Some of the criticisms he found associated with evaluating teaching were considerations 1) that it could lead to unhealthy competition among faculty members, 2) that student inputs would not be valid since they could be misleading or inadequate, 3) that the teaching process could become rather stereotyped and 4) that no one knows what constitutes effective teaching. The tendency has been to recognize the resistance and then move along with the development of methods to do the evaluation. It has not been sufficiently appreciated that this resistance will make it difficult for an evaluation system to be accepted and effectively employed. A well-designed system could attenuate some of this resistance. Current proposals, however, are rather narrowly conceived and inadequate to gain wide faculty support. Rather than minimize the threat to self-esteem and job security, they will more than likely only increase it.

There are several factors associated with current methods and assumptions which contribute to this problem. In the following paragraphs, I will describe these factors, specify their implications for the evaluation process and suggest some ways that the problems might be overcome.

HE 003 717

To evaluate teaching implies that one knows the relationship between teaching and learning. Unfortunately, it is not clear just what the relationship really is. Even so, evaluation systems are established without considering this issue in depth. In looking over some of the questionnaire data from other colleges, it would seem that the designers have implicitly answered the question. A casual glance at any questionnaire designed to evaluate a teacher reveals many familiar queries about the behavior of the teacher—teacher explains clearly, is friendly toward students, posseses self-confidence. There are two problems with the approach implied by such items. First, the activity of students in the teaching-learning process is ignored. Teacher interpersonal behaviors and methods are only effective to the point that students are willing and able to use them. What the student *really* needs from the teacher, in order to learn, is not emphasized in this approach. With regard to learning the content of a course, most of this takes place in the privacy of the student's study area. The usefulness to the student of such things as the teacher's sense of humor, self-confidence, greeting students outside of class and other similar targets of evaluation instruments is not clear. If content acquisition is one of the major goals of education, what relevance do these items have to how well a teacher accomplishes that goal?

A second, related problem with this approach is that it assumes ideal sets of teacher interpersonal behaviors and classroom methods. If this were true, one would expect that some teacher interpersonal behaviors and classroom methods would be more effective than others in this process. Otherwise, why ask questions about how well teachers do things? Unfortunately, the research literature offers little support for making that assumption. With regard to content acquisition, it is becoming clear that a wide variety of teaching methods (lecture, lecture discussion, discussion) lead to approximately the same amount of content acquisition, as R. Dubin and T. C. Taveggia argue in *The Teaching-Learning Paradox*. Similarly, there is no clear evidence that a teacher's qualities such as availability to meet with students, possessing a friendly manner, showing self-confidence or a sense of humor are positively related to various amounts of content acquisition. While students may be more satisfied or interested in class when teachers exhibit such behaviors, the effect that has on their learning is not clear. To date, research on the relationship between rating scale categories and content learning has failed to show that there is any substantial relationship.

One might argue, of course, that although there are few differences in content acquisition, different classroom methods and teacher interpersonal behaviors may affect the student's satisfaction with a course or the learning of non-content skills (e.g., positive societal values, collaborative interpersonal skills, ability to use resources). There are several issues to take with this thinking. Granted that distinguishable methods might lead to changes in non-content behaviors of students, it is not clear what methods or teacher interpersonal behaviors work best. Furthermore, not every teacher is interested in establishing goals in addition to content-learning for his or her classes. To evaluate all teachers on this basis would be an obvious injustice to many.

Judging, then, from these considerations, an evaluation system should not make *a priori* judgments about ideal teacher interpersonal behaviors and classroom methods. Teachers have known intuitively for a long time that there are no such things. It is not surprising that they resist instruments that imply that such behaviors exist. A better way to initiate the evaluation of teaching is to ask teachers what their goals are for a class. Presumably, what they do in the classroom is related to their goals. One aspect of the evaluation would be how well the classroom environment is related to the student's ability to obtain content and/or non-content goals. No ideal set of teacher behaviors or goals are implied in this approach. Since different instructors have different goals, a variety of interpersonal behaviors and methods would be expected. The evaluation interest is in whether the instructor was able to accomplish stated goals and which of his or her activities need to be improved.

To state as precisely as possible the goals for a course, one could begin, for example, with general goals of content acquisition and the developing of interpersonal collaborative skills in students. For example, "Students should be able to list the important dates and events associated with a given period of history." Or, "The student should be able to write a satisfactory term paper in collaboration with two other students." Students would then be polled to see how well things like lectures, outside readings and small group meetings allowed these particular goals to be obtained. The real test of the effectiveness of a teacher's behavior is how the things he or she did were seen as useful contributions to the student's experience with the course.

One advantage here is that instructors are free to specify the type of classroom environment they want to establish and be evaluated with regard to how well their methods allowed their goals to be met. If one goal was to show students that teachers are human beings, then an instructor might do this by meeting with students informally, or appearing friendly. However, no assumptions are made "across the board" as to the desirability of any one behavior for every instructor. Faculty members should feel less threatened under this approach since they are not asked to conform to so-called "ideal" behaviors. Furthermore, the use of student feedback could serve as a basis for modifying goals and/or methods to improve the classroom environment. This process could only lead to employment of a wider variety of teacher interpersonal behaviors and methods. Any subsequent changes in classroom activities would be the result of an objective examination of the methods previously employed. Surely this is better than the subjective approach most often used today. Finally, this approach forces the evaluator to consider each class situation as a unique entity. While this means more work for the evaluator, such a system should contribute less to faculty resistance.

Classroom-related behaviors are certainly only one measure of a teacher's effectiveness. One should consider the instructor's influence once the student leaves the classroom. Some areas that should be considered are how the student perceives a given teacher as influencing a) student's major field of study, b) types of course electives, c) decision to do post-graduate study, d) decision about what job to take and e) how well the instructor prepared the student to work or study in a field. Obviously, the above are not the only things one could consider. They are merely examples of some of the ways an instructor's influence might appear outside of class. The important point to recognize is that effectiveness criteria should not be locked into classroom-related behaviors.

How a teacher is perceived by his peers is an equally important part of an evaluation system. Colleagues can provide insights related to what they feel an instructor's academic and personal strengths and weaknesses are. Since such opinions have implications for how a teacher relates to colleagues and students, these thoughts should be brought to the surface. Few instructors know where they stand in relation to their colleagues and how their behaviors are judged by them. For the instructor who is interested in improvement this can be valuable information.

Finally, an instructor's professional activities should be counted in an evaluation. Factors like research, professional writing, involvement in professional organizations and meetings and involvement in community affairs are important. It must be noted that these things should be compared with the values of the educational institution. Conflicts can occur when the instructor's values and those of the institution are at variance (e.g., on the importance of research). One function an evaluation could serve is to show instructors where the discrepancies exist early in their careers. I suspect that some of the surprise at not being promoted would disappear if such discrepancies were made explicit.

It should be clear at this point that I regard classroom activities as only one area that should be considered in developing an evaluation system. Many recent attempts at evaluation have not adequately incorporated this thinking. Consequently, and understandably, faculty members are cool toward evaluation processes that rely exclusively on classroom-related behaviors.

But the most neglected problem is just how an evaluation system ought to be run and what methods of data collection might be used successfully. I have three major suggestions to make: First, questionnaires and personal interviews, though useful techniques for obtaining data, cannot be designed intuitively. Professional consultation should be sought in the design and implementation of these methods if they are to be used by someone not trained in how to develop them. To measure a teacher's abilities, we want the most reliable and valid procedures available. While an informal approach might appear to be less threatening than a more rigorous one, that is exactly the approach most likely to suffer from deficiencies in method. A familiar problem is the one in which faculty members recognize those deficiencies and then, appropriately, reject the data. Anything that is worth measuring is worth measuring well.

As a second point, questions about faculty behavior should be written carefully. A little-known phenomenon is a student's tendency *not to rate the actual behavior of the faculty member*. Instead the rating is made on how much that behavior deviates from a reference point that the student employs. For example, the instructor may be considered good or poor with reference to how some "ideal" teacher behaves or in terms of the "average ability of the student's other instructors." In other words, the rating tends to reflect the discrepancy between the student's refer-

ence point and his impression of the instructor he is evaluating. With traditional rating procedures, these reference points are not assessed directly. The extent to which low ratings (or high ones) reflect the same degree of discrepancy between the reference point and the actual behavior depends upon the assumptions that all students use the same reference point and have the same level of expectations for that reference point. That is, no problem would occur if all students used the average ability of all their instructors as a reference point and rated this ability as "good" on a given question. The rating of the actual behavior for that category (e.g., organization of lectures) would represent the same discrepancy for all students. Unfortunately, research just completed by this author shows that students use more than one reference point and that the reference points differ by student and the types of questions asked. Since the reference-point value systems of students are different, individual evaluations on a rating scale are basically incomparable; they cannot be legitimately combined. This consideration questions the appropriateness of all traditional rating procedures that are so often employed in teacher evaluation questionnaires.

There are at least two ways that this problem might be minimized. One way is to assess the student's reference points with regard to the teacher's behavioral categories under consideration. Having this information would help to make reasonable interpretations of the behavior ratings possible. Another course is to look specifically at how well the teacher's behaviors and methods helped students meet the goals of the course. Taking this approach, one is less interested in rating a set of teacher behaviors in terms of "good," "poor," and so forth. For example, evaluative questions touching on whether the teacher was "friendly" or whether he or she assigned a "large amount" of reading would not be asked in terms of "good" or "poor." Rather the emphasis would be on seeking specific ways that the "friendly" behavior and reading assignments helped or hindered the student's learning. The orientation of the questions would be different since they would concentrate on the specific effects of teacher interpersonal behaviors and methods on the student's performance. A further payoff of this plan is that the type of information collected is more useful to the teacher since it can be used to improve specific behaviors. A rating of "good" on friendliness might be useful because students were not afraid to ask for clarifications of content after class.

How long an evaluation should last and who should conduct it is the third major issue demanding attention. There is a tendency in the literature to assume that students or interested faculty can write a questionnaire, administer it, publish the data and then regard the job as complete. If my argument in this article is at all legitimate, it should be apparent that a quick and dirty job would yield rather poor data. To assess a teacher with regard to just some of the considerations suggested here will take time. I doubt that an adequate assessment of what positive and negative effects an instructor is having could be accomplished in less than one year. This assumes that one is interested in obtaining information, giving the instructor feedback, and then reassessing to see whether positive changes have occurred in the instructor's behavior. This is particularly important when promotion and tenure considerations are tied in with the evaluation effort.

Who should conduct the evaluation is something that likewise must be taken seriously. Some possibilities are an independent evaluation agency, interested students, faculty and administrators in a college, or perhaps a committee composed of junior and senior members of a department. Who does it depends upon what purpose the evaluation is to serve. Again, this is particularly important when promotion and tenure are involved. The important points are that a group ought to be made up of people whom faculty members can trust and that competent evaluation guidelines and goals be established. The written expression of guidelines and goals should be done in conjunction with competent professional advice and with the use of inputs from as many relevant sources as possible. Before the evaluation is undertaken, the guidelines and goals should be communicated to the faculty members who will be evaluated.

I hope that some of the conceptual, methodological and operational suggestions made in this article will help to attenuate the resistance towards an evaluation of teaching effectiveness. Furthermore, it is my belief that asking questions about how *specific* teacher behaviors were or were not useful inputs to students for meeting the course goals should give the teacher information for changing the classroom environment intelligently. Finally, I would suggest that the data collected be used initially for feedback purposes. By this I mean that before evaluation data is used for promotion and tenure considerations, the instructor *should be given the opportunity to improve. How much* he improved is not the issue for promotion and tenure.

Rather, those who make the decision should ask the following questions:

First, did the instructor make an attempt to improve what he or she was doing? Professionals should be concerned with improving what they do. Instructors who are not interested in improving their techniques are not worth having on a faculty.

Second, are the instructor's values with regard to things like professional research, writing and community service compatible with the institution's values? Feedback to the instructor would make this clear. If the institution considers any or all of these things important, then the instructor should be told what is expected. How well the faculty person conforms to these values would be a promotion and tenure consideration.

Third, do colleagues perceive the instructor as a valuable member of the faculty? Presumably, instructors would be given feedback consisting of how colleagues perceive them as professionals. This could help them to work on relationships and to modify behaviors in ways that are less antagonizing to colleagues. Interest here would not be so much with interpersonal habits as they would be with how the teacher is perceived as sharing the administrative load (e.g., advising students, serving on committees, participating in faculty meetings, etc.).

Fourth, have students reported that what they learned from the course, in class and out, helped them in decisions about what to major in, post-graduate study or on the job? Teachers who are considered effective should be those seen as contributing something of use to the student beyond the course itself. If deficient in this area, a concerned instructor would try to change what he or she is doing.

A broadly-conceived evaluation system should give those who make promotion and tenure decisions the appropriate information. How various inputs should be weighed is another issue. Assigning differential weights to each category (or to any others that are developed) is really not useful and will needlessly raise value issues. Professionals should be concerned with improving their professional behaviors. Based on changes in the inputs obtained from a broadly-conceived evaluation enterprise and considered with respect to clearly-expressed goals, it will become quite clear who the professionals are. The important considerations in who to promote and/or give tenure to are whether a faculty member is honestly concerned with improvement and can be objectively seen as having taken steps in this direction. Only those concerned about improving and who try deserve the title of teacher.

—Anthony F. Grasha

Plan For Individual Learning

At least six faculty members will participate, through unpaid overload, in a plan to develop more self-directed learning in the College of Community Services, University of Cincinnati. About thirty-five students will be enrolled in this program, which is to begin in September, 1972.

Several ideas underlie this curriculum option. First, interested students will have the opportunity to develop, with the help of the faculty, programs which are suited to their own needs and interests. Second, students and faculty members will be able to give greater attention to evaluating learning as distinct from working for or assigning course grades. Third, through participation in core projects, students and teachers will be able to develop learning experiences which are not clearly within the provinces of separate departments and which integrate a broad range of professional skills (e.g., community action, research, interpersonal and organizational competencies, professional ethics and responsibilities, and holistic problem-solving). And finally, the college will attempt to help its alumni and professionals in the Cincinnati area in continuing learning and in efforts to become innovators in the human service professions.

At least one-third of each student's program will be devoted to participation in a core project. Students and faculty members participating in the program will join one of several project teams for the academic year. (Other students, teachers, and members of the larger community may work with these teams from time to time; for example, some faculty members may be consultants for certain skills or issues.) Each proj-